Desalination and Water Treatment
www.deswater.com

doi: 10.5004/dwt.2017.20911

84 (2017) 93–101
July

Geostatistical cokriging and multivariate statistical methods to evaluate groundwater salinization in Faisalabad, Pakistan

Maqsood Ahmad[a], Sohail Chand[a,*], H.M. Rafique[b]

[a]*College of Statistical and Actuarial Sciences, University of the Punjab, Lahore, Pakistan, emails: sohail.stat@gmail.com (S. Chand), maqsoodzain5@gmail.com (M. Ahmad)*
[b]*Department of Physics, University of the Punjab, Lahore, Pakistan, email: hafizrafique42b1@hotmail.com (H.M. Rafique)*

Received 29 November 2016; Accepted 15 May 2017

ABSTRACT

In the present study, multivariate techniques and geostatistical cokriging were used to look into the groundwater salinization of district Faisalabad. The groundwater condition of district Faisalabad has become miserable because of the rapid increase in population, industrial wastes, and agrochemical application. As a result, majority of the people do not have access to pure drinking water and, consequently, polluted water is causing many deaths per year. A number of 220 water samples based on instructions of World Health Organization (WHO) were taken from four main sources such as hand pump, injector pump, tube well, and water supply. All samples were tested for 12 water quality parameters and summary statistics were calculated to compare the water quality parameters with WHO permissible limits. Initially, correlation matrix was constructed to evaluate the most significant parameters and later on used principal component analysis (PCA) to select those parameters causing maximum variation. Dendrogram based on cluster analysis conveyed same sort of information as delivered by PCA. First, factor of PCA contributed 41.6% of total variation. Six water quality parameters such as sulfate, calcium, total dissolved solids, sodium, chloride, and magnesium were found to be most alarming because all of these have factor loadings greater than 75%. Cross-variogram based on cokriging showed spatial dependence as well as positive pairwise spatial correlation among all parameters. The prediction maps highlighted the most dangerous and health hazard areas; therefore, may be very helpful for water management agencies to target those high-risk areas. It was found that the area with east latitude 31.0°–31.4° and north longitude 72.8°–73.2° is most alarming zone.

Keywords: Spatial dependence; Health effects; Drinking water; Dendrogram; Cross-variography

1. Introduction

Groundwater is a main source of pure water and freshwater in arid areas to fulfil the agricultural, industrial, and domestic requirements [1]. However, an excessive usage of groundwater aquifers shrinks the water quality. Furthermore, inappropriate removal of waste materials, industrial waste in water bodies, food residue, and excessive usage of agrochemicals in agriculture contaminate water. Waterborne diseases are considered as a major reason of children's death in developing countries [2]. Two billion people round the globe use groundwater for drinking purposes [3]. In Pakistan, people get drinking water through several resources like tube well, hand pump, house water supply, and injector pumps. Since, little attention has been given to improve the drinking water quality; therefore, water obtained from these resources is often contaminated [4]. Water supply is mostly irregular and waterborne diseases such as typhoid, hepatitis, diarrhea, and stomach inflammation are common in Pakistan [5]. In city areas, water sanitary condition is also unsatisfactory [6].

In Pakistan, about 30% of diseases and 40% of deaths are due to contaminated and impure drinking water [5]. Since, every fifth inhabitant of Pakistan is suffering from waterborne infections; therefore, 0.1 million deaths (with 250,000

* Corresponding author.

children) occur each year [7]. It is need of the hour to identify the factors which cause this severe groundwater contamination. Ullah et al. [8] investigated the factors which are contaminating the groundwater near Kabul river Peshawar, Pakistan. They suggested to install wastewater treatment industries near Peshawar to cope with the most contaminated water. Waseem et al. [9] reviewed various factors that cause pollution in water, soil, and vegetables. They discussed pollution status in major areas of Pakistan. Iqbal and Khera [10] stated that copper and lead are introduced into water due to industrial waste and other human activities. Similarly, Iqbal et al. [11] investigated nine water quality parameters including pH, sulfate, chloride, and total dissolved solids (TDS) after treatment of the contaminated wastewater.

Recently, many statistical techniques such as factor analysis, cluster analysis (CA), principal component analysis (PCA), and discriminant analysis have been used to predict the distribution of groundwater quality parameters due to an increase in physical and chemical variables in groundwater [1]. These techniques not only lead to effective learning of water quality characteristics but also provide reliable solutions to enhance the quality of water [12–15]. Geostatistics suggests various methods to model and predict the spatially varying groundwater parameters and pollutant concentration [16]. Mainly, geostatistical analysis consists of variogram modelling to assess the correlation structure, kriging, cross-validation, and spatial mapping.

Kriging has been extensively used to evaluate spatial variability of groundwater quality parameters [17,18]. Nas [17] used ordinary kriging to assess the spatial variation of water quality parameters in Turkey. Ahmad and Chand [18] compared ordinary and Bayesian kriging to evaluate the spatial distribution of TDS level in groundwater of tehsil Jampur, Pakistan. They used Box–Cox transformation to normalize the positively skewed spatial variable and used Matern covariance model to evaluate the correlation structure. Ahmad et al. [19] showed the outperformance of Bayesian kriging while predicting the sulfate concentration in groundwater of Jampur, Pakistan. They assessed the spatial autocorrelation among sampled observations using variogram envelop [20].

Pozdnyakova and Zhang [21] remarked that kriging is used to assess the spatial distribution of a sampled variable while cokriging is a multivariate geostatistical method used to improve the estimation procedure of sampled variables as it takes into account the cross-correlation with the better sampled variable. Triki et al. [1] compared cokriging with ordinary kriging to assess the groundwater salinization of eastern Tunisia and demonstrated the effectiveness of cokriging through cross-validation. They evaluated the most significant water quality parameters by PCA, CA, and multivariate correlation matrix and took TDS as auxiliary variable while Na^+, Cl^-, SO_4^{2-}, and sodium absorption ratio (SAR) as primary variables and showed the prediction results by contour maps. Mehrjardi et al. [22] predicted the groundwater distribution of Yazd-Ardakan plain, Iran, using inverse distance weighting, ordinary kriging, and cokriging. They used spherical and exponential covariance structures to assess the underlying correlation among water quality parameters. Their results supported the use of cokriging geostatistical method.

In this paper, our main objective is to illustrate the use of geostatistical method combined with multivariate statistical method in order to evaluate the quality of groundwater in Faisalabad. First, we used multivariate correlation matrix, PCA, and CA to identify the most significant groundwater quality parameters. Second, an unbiased and efficient multivariate geostatistical cokriging technique along with suitable spatial correlation structure is used to improve the estimation of groundwater salinization. This also leads to the evaluation of the spatial distribution of highly contaminated groundwater quality parameters. Finally, we draw the prediction maps of most significant water quality parameters.

2. Materials and methods

Since water pollution is mainly caused by the toxic chemicals; therefore, 12 water physiochemical parameters such as TDS, sulfate SO_4^{2-}), nitrate NO_3^- , potassium (K^+), sodium (Na^+), magnesium (Mg^{2+}), calcium (Ca^{2+}), bicarbonate HCO_3^-, pH, chlorides (Cl^-), fluoride (F^-), and SAR have been studied. Among these parameters, TDS and pH are physical parameters while remaining are chemical parameters. TDS consists of organic matters and inorganic salts dissolved in drinking water. Most common inorganic salts are Na^+, Ca^{2+}, K^+, Mg^{2+}, Cl^-, HCO_3^-, SO_4^{2-}, and NO_3^-. It is noteworthy that the chemical parameter SAR illustrates the water suitability for irrigation purpose. As declared by Triki et al. [1], SAR is expressed as:

$$\mathrm{SAR} = \frac{\mathrm{Na}^{+}}{\sqrt{\dfrac{(\mathrm{Mg}^{2+} + \mathrm{Ca}^{2+})}{2}}} \tag{1}$$

where Na^+ is sodium concentration, Mg^{2+} is magnesium concentration, and Ca^{2+} is calcium concentration. According to World Health Organization, if level of SAR is ≤13 then water is good for irrigation purposes otherwise soil will become sodic and result in decreased agricultural yield.

2.1. Study area of the hydrogeochemical data

The study area, Faisalabad is situated in Punjab province in eastern Pakistan at east latitude 31.4292° N and north longitude 73.0789° E. It is the most famous and populated city after Karachi and Lahore. According to Pakistan Demographics Profile [23], population of Faisalabad has reached 3.567 million. It is known as Manchester of Pakistan for its textile industry and it contributes about 20% in annual gross domestic product of Pakistan. It is well-known industrial city because it lies at center of many cities connected through highways, railway tracks, and by air transportation. Rapid growth in population and increase in number of industries has increased the level of wastewater extensively which is consequently contaminating the resources of freshwater and surrounding environment [7]. To observe the quality of groundwater in Faisalabad, a survey was conducted by Pakistan Council of Research in Water Resources (PCRWR), based on 220 water samples collected from tube well, hand pump, house water supply, and injector pumps (see Fig. 1). Guidelines provided by World Health Organization (WHO) [24] were followed during the survey.

2.2. Multivariate statistics

Multivariate statistical techniques are effective tools and used extensively for analysis of water physicochemical parameters [25] and to study the groundwater contamination [26]. The sample correlation coefficient between *i*th and *j*th variables is illustrated as:

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}} = \frac{s_{ij}}{s_i s_j} \tag{2}$$

The correlation matrix for a sample data corresponds to covariance matrix of same data with correlation as a substitute of covariances.

$$\mathrm{R} = r_{ij} = \begin{bmatrix} 1 & r_{12}\cdots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{n1} & r_{n2}\cdots & 1 \end{bmatrix} \tag{3}$$

PCA is often performed on groundwater parameters to learn the interrelations among parameters and to shrink the number of water quality parameters. It is a famous multivariate statistical data analysis technique due to its algebraic simplicity and straightforward interpretation. In this method, a linear transformation is defined in such a way that it changes the correlated variables into uncorrelated orthogonal factors. CA is also a multivariate statistical method where dendrogram is drawn in order to find the most collinear and associated clusters. Usually, the results of CA confirm the findings of PCA [1].

2.3. Theory of cokriging

Kriging is a technique that enables prediction of a spatial process based on a weighted average of the observations. In case of an intrinsically stationary process with constant unknown mean, we use the ordinary kriging method. Cokriging is a geostatistical technique that deals multivariate data to estimate the spatial dependence structure of physically correlated variables [27]. It is a modification of geostatistical ordinary kriging method. In cokriging estimation, an undersampled variable is taken as a primary variable while the variables sampled at low cost are taken as secondary variables. Secondary variables are used to establish cross-correlation with primary variables. Thus, this technique is suitable for prediction of an important variable as a function of other variable. The best linear unbiased estimate of Y at any unobserved location x_0 can be mathematically expressed as:

$$\hat{Y}(x_0) = \Sigma_{i=1}^{n} \lambda_i \, Y(x_i) + \Sigma_{k=1}^{m} \omega_k Z(x_k) \tag{4}$$

where $Y(x_i)$ is the observed values of primary variable Y at locations x_i; and $i = 1, 2, \ldots, n$ and $Z(x_k)$ are the observed values of secondary variable Z at locations x_k, $k = 1, 2, \ldots, m$. In Eq. (4), λ_i and ω_k are weights of cokriging technique. These weights are chosen in such a way that the estimate remain unbiased with least variance [1]. Cokriging has also the advantage of assessing the pairwise cross-correlation structure among regionalized variables. This is possible using cross-variogram which is an integral part of cokriging analysis. Let $Y(x_i)$ and $Z(x_k)$ are two random variables, the cross-variogram under the second order stationarity is given as:

$$\gamma_{ik}(h) = \frac{1}{2} E\{[Y(x_i + h) - Y(x_i)][Z(x_k + h) - Z(x_k)]\} \tag{5}$$

This is also known as the linear model of regionalization using variogram. A detailed description of cokriging system has been illustrated by Subyani and Al-Dakheel [27].

3. Results and discussion

3.1. Exploratory data analysis

We used R statistical software [28] to analyze the hydrological data. Initially, basic statistics of all water quality parameters along with normality test and permissible limits are given in Table 1. It can be seen that all water quality parameters are violating the permissible limits as described by WHO [24]. The coefficient of variation (CV) represents the extent of variation of each parameter concentration. All samples yield moderate to high spatial variation among observed samples of the water quality parameters. The high CV for Cl^-, NO_3^-, K^+, and Ca^{2+} reflects the high spatial variation in groundwater of Faisalabad.

In order to evaluate the most significant and correlated water quality parameters, correlation matrix is displayed visually (Fig. 2) for all 12 physiochemical parameters. Results based on Pearson's correlation show that TDS have high correlation ($r \geq 0.8$) with Cl^-, Mg^{2+}, and SO_4^{2-}. Similarly, Na^+ shows pairwise high association with TDS, Cl^-, Mg^{2+}, and SO_4^{2-}. Mg^{2+} shows positive correlation with TDS, Cl^-, and Na^+. It is also noted that SO_4^{2-} exhibits highly positive association with four water quality parameters like Mg^{2+}, Cl^-, TDS, and Na^+. Hence, a significant positive correlation ($r \geq 0.7$) between TDS, Mg^{2+}, Ca^{2+}, Na^+, Cl^-, and SO_4^{2-} is evidenced.

3.2. Principal component analysis

PCA is used to establish the association between water quality parameters and to assess the correlation between these parameters. In this study, 12 water quality parameters: TDS, SO_4^{2-}, NO_3^-, K^+, Na^+, Mg^{2+}, Ca^{2+}, HCO_3^-, pH, Cl^-, F^-, and SAR have been used for PCA. Results of PCA (Table 2) revealed that first three factors are able to show the hydrogeochemical procedures of the groundwater without any significant loss of information. PCA generated three factors which collectively account for 76% of total variation. First factor account for 41.6% of total variation which is associated with high loading of TDS, Mg^{2+}, Ca^{2+}, Na^+, Cl^-, and SO_4^{2-}. These six parameters constitute highly strong loading (≥0.7). Further, in Factor 1, all loadings are positive except pH (–0.015). Therefore, it is considered as an important factor which mainly affecting the groundwater quality. Factor 2 account for 25.5% of total variation. In Factor 2, positive loadings are associated only with Mg^{2+} and Ca^{2+} while all other parameters indicate negative loadings. The Factor 3 explains only 8.9% of total variation which is mainly related to NO_3^- having high negative loading (–0.93).

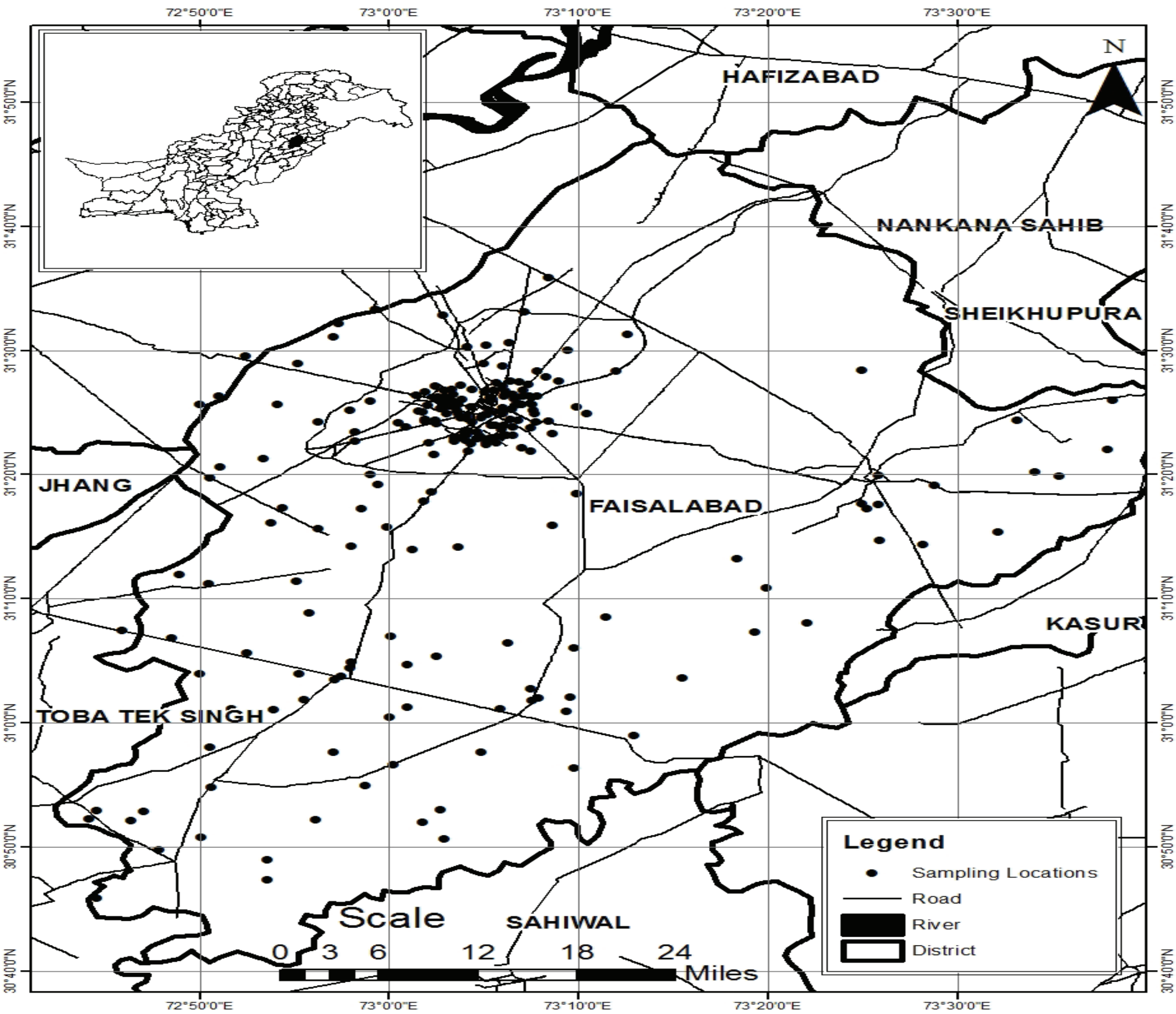

Fig. 1. Location map displaying the 220 monitored sites observed for groundwater in district Faisalabad.

Table 1
Descriptive statistics of physiochemical parameters in groundwater of district Faisalabad

Variable	Units	Minimum	Maximum	Mean	SD	CV	AD normality test		Permissible WHO limit
							A^2	p Value	
TDS	mg/L	104	9,804	1,961.8	1,560.5	79.54	8.68	0.0011	≤1,000
Ca^{2+}	mg/L	4	1,100	81.27	83.31	102.50	20.09	0.0002	–
Mg^{2+}	mg/L	0	287	62.402	54.134	86.75	11.71	0.0021	≤150
HCO_3^-	mg/L	38	1,210	448.81	199.10	44.36	2.08	0.0021	–
pH	–	1.4	70	9.2464	5.636	60.96	8.47	0.0013	6.5–8.5
Cl^-	mg/L	11	4,255	500	593.05	118.61	17.13	0.0019	≤250
K^+	mg/L	0	182	18.630	19.98	107.25	15.30	0.0000	≤12
Na^+	mg/L	3	2,000	607	513.99	84.68	7.48	0.0030	≤200
SO_4^{2-}	mg/L	12	2,416	588.14	518.40	88.14	10.20	0.0000	≤500
NO_3^-	mg/L	0	21.20	3.008	3.33	110.98	12.92	0.0002	≤50
F^-	mg/L	0.02	4.22	0.984	0.5587	56.78	3.05	0.0001	≤1.5
SAR	–	0.677	260.768	75.573	58.95	78	4.91	0.0032	≤13

Note: AD, Anderson–Darling.

3.3. Cluster analysis

CA is also performed on the set of 220 samples over 12 water quality parameters. The main objective of performing the CA is to evaluate association between water parameters. Dendrogram (Fig. 3) has grouped all 12 water parameters into 3 significant clusters. First cluster consists of TDS, Mg^{2+}, Ca^{2+}, Na^{+}, Cl^{-}, SO_4^{2-}, and K^{+}. Since the loading of potassium (K^{+}) is <0.75 (weak association). Therefore, we consider only six significant factors for prediction purposes. Overall, the results of CA confirm the classification of PCA and correlation matrix. It is also noteworthy that nitrate concentration is uncorrelated with all remaining water quality parameters and hence fall in a separate cluster.

3.4. Geostatistical estimation

Geostatistics suggests many spatial estimation techniques which are used for interpolation of any observed variable at unsampled locations. These procedures are known as kriging that provide us the measures of accuracy in the form of estimated variance [29]. Many kriging techniques like simple kriging, ordinary kriging, universal kriging, block kriging, cokriging, lognormal kriging, and indicator kriging are famous minimum variance interpolation techniques.

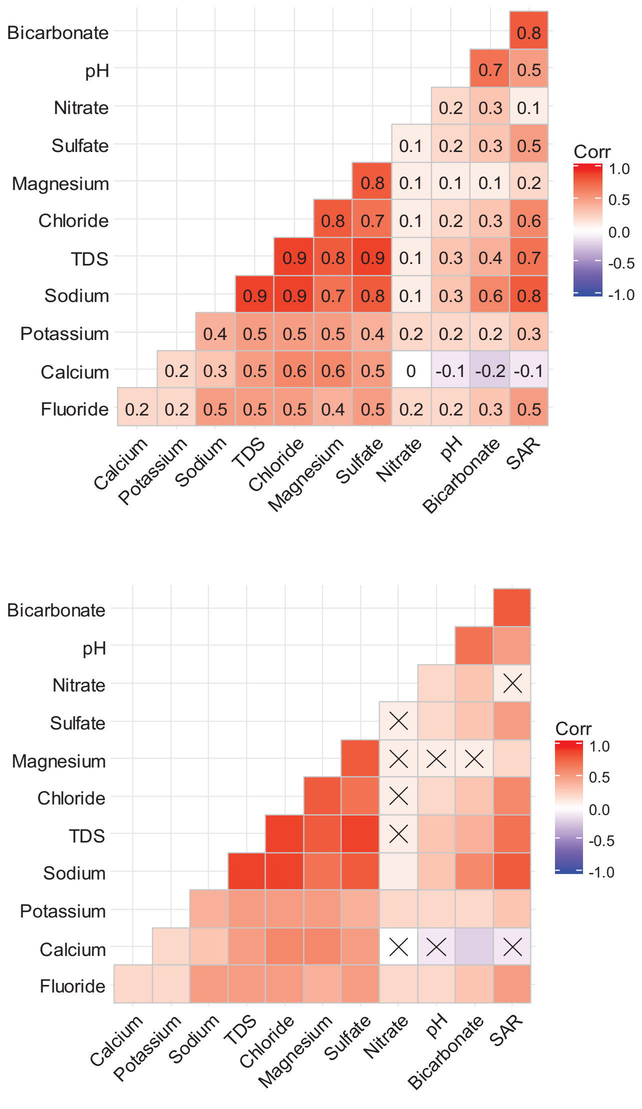

Fig. 2. Multivariate correlation matrix displayed graphically (top panel) and variables having insignificant correlation (bottom panel).

3.4.1. Coregionalization

An important step in cokriging is to build a proper model assessing the dependency among underlying variables [1] which is referred as coregionalization and estimated using cross-variogram. We draw the cross-variogram using *gstat* and *vgm* functions of *gstat* package [30] of R statistical software [28]. All these direct semivariograms generally indicate some spatial structure and most of them have moderately large component of nugget effect. These semivariograms show that the ratio, nugget/total variation, of TDS, Ma^{2+}, Ca^{2+}, Na^{+}, Cl^{-}, and SO_4^{2-} was between 15% and 60% which indicates the spatial dependence of these variables.

Furthermore, Fig. 4 shows that many cross-variogram models are close to the maximum correlation (≈0.9) except the cases TDS–Ca^{2+}, Ca^{2+}–Na^{+}, Ca^{2+}–Cl, and Ca^{2+}–SO_4^{2-}. This reduction in the spatial cross-correlation is due to their small correlations TDS–Ca^{2+} (0.5), Ca–Na^{+} (0.3), Ca^{2+}–Cl^{-} (0.6), and Ca^{2+}–SO_4^{2-} (0.5). In spite of these, no pair shows negative spatial cross-correlation.

3.4.2. Ordinary kriging prediction of target variable

In this section, we use ordinary kriging technique for the spatial prediction of the target variable, TDS without using covariables. Initially, we estimated the three parameters of the spherical variogram model ($\sigma^2 = 2.35$, $\phi = 10.12$, $\tau^2 = 5.12$) using *eyefit* command and later confirmed the parameters estimation using ordinary least square method. Afterwards, we used *krige.control* function of geoR Package [31] to obtain

Table 2
Pattern of rotated factors using varimax rotation of 12 water quality parameters along with cumulative % variance

Variable	Factor 1	Factor 2	Factor 3
TDS	0.907	–0.383	0.001
Ca^{2+}	0.757	0.367	–0.050
Mg^{2+}	0.913	0.025	–0.125
HCO_3^{-}	0.092	–0.890	–0.188
pH	–0.015	–0.738	–0.173
Cl^{-}	0.895	–0.271	0.019
K^{+}	0.529	–0.142	–0.325
Na^{+}	0.783	–0.573	0.031
SO_4^{2-}	0.851	–0.263	0.012
NO_3^{-}	0.055	–0.148	–0.932
F^{-}	0.455	–0.441	–0.036
SAR	0.363	–0.859	0.104
% Total variance	41.6	25.5	8.9
Cumulative % variance	41.6	67.1	76

the predicted values. At the end, we use the leave-one-out cross-validation to calculate the prediction errors (Table 4).

3.4.3. Map-based cokriging analysis

In order to predict the spatial autocorrelation structure of groundwater quality parameters and to assess the groundwater salinization in Faisalabad, the geostatistical cokriging is applied. Results based on multivariate correlation matrix, PCA, and CA suggest highly significant variables. Significant pairwise associations were found among TDS, Mg^{2+}, Ca^{2+}, Na^+, Cl^-, and SO_4^{2-}. Further, all these variables crossed the permissible limit (Table 1). Therefore, these variables have major contribution in contaminating the groundwater. To carry out geostatistical cokriging, TDS is considered as auxiliary variable while Mg^{2+}, Ca^{2+}, Na^+, Cl^-, and SO_4^{2-} are taken as primary variables. Since, all geostatistical prediction methods are based on the assumption of normality of observations; therefore, Anderson–Darling normality test (Table 1) revealed that all of these variables are positively skewed. Box–Cox transformation [32] used by Ahmad and Chand [18] and Ahmad et al. [19] is carried out by using CAR package [33] of R statistical software [28] to normalize these variables.

An important step in geostatistical prediction is to assess the spatial autocorrelation using spatial anisotropy or isotropy. A spatial structure is isotropic when the pattern of the spatial correlation changes due to the change in the direction of orientation of pairs of locations [19]. It was found that the

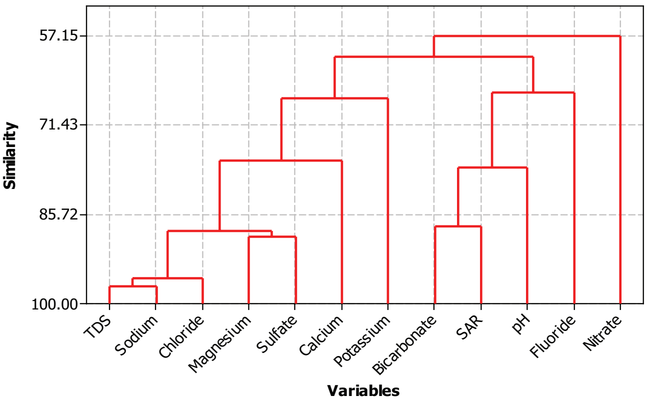

Fig. 3. Dendrogram based on cluster analysis showing the relations among 12 water quality parameters.

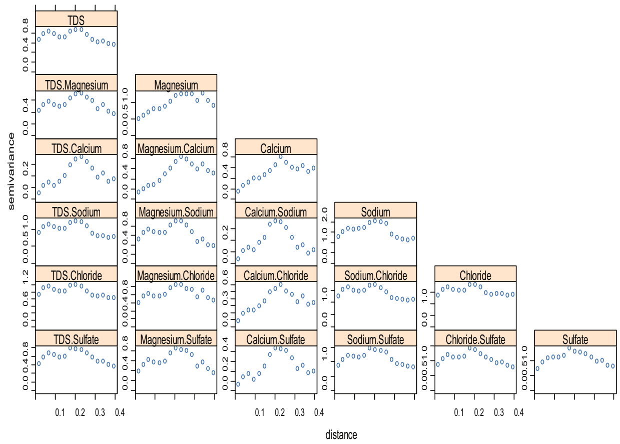

Fig. 4. Cross-variography of groundwater quality parameters of district Faisalabad, Pakistan to assess the cross-correlation structure.

Table 3
Parameters of best fitted variogram models of significant physiochemical parameters

Groundwater parameters	Variogram model	Estimation method	Sill (σ^2)	Range (ϕ)	Nugget (τ^2)	τ^2/σ^2	RMSPE
TDS	Spherical	REML	2.35	10.25	5.12	2.17	0.6587
Mg^{2+}	Exponential	REML	3.14	11.24	3.45	1.13	0.2354
Ca^{2+}	Exponential	MLE	4.53	10.00	2.56	0.51	0.4789
Na^+	Matern	REML	2.22	9.89	3.25	1.03	0.9875
Cl^-	Exponential	REML	2.12	8.65	6.47	3.12	0.8425
SO_4^{2-}	Matern	MLE	1.99	6.25	8.57	4.39	0.7746

Note: RMSPE: root mean square prediction error.

Table 4
Cross-validation statistics for cokriging and ordinary kriging

Parameter	Ordinary kriging				Cokriging			
	MSSE	MSE	ME	R^2	MSSE	MSE	ME	R^2
TDS	1.09	2,350	–5.78	0.19	1.01	1,504	–2.3	0.91
Magnesium	0.91	6,113	–10.96	0.41	1.1	1,384	2.3	0.89
Calcium	1.2	1,592	–17.59	0.35	1.14	838	0.66	0.84
Sodium	0.96	10.6	0.04	0.11	0.97	0.65	0.03	0.94
Chloride	1.2	4.49	–0.09	0.51	0.97	0.51	0.01	0.93
Sulfate	1.25	3.32	0.05	0.34	0.99	0.49	0.03	0.85

structure of spatial autocorrelation of response variable was approximately independent from direction.

As revealed by PCA, concentrations of Mg^{2+}, Cl^{-}, Na^{+}, Ca^{2+}, and SO_4^{2-} were taken as primary variables while TDS is taken as auxiliary variable. Geostatistical cokriging basically studies the spatial continuity and dependence among variables of interest. Initially, variogram of all variables are computed and parameters (sill, range, and nugget) are estimated by maximum likelihood estimator (MLE) and restricted maximum likelihood (REML) (Table 3). Four water quality parameters support the REML estimation technique with spherical covariance function. We assess the property of spatial dependence [20] by the ratio nugget/sill. It is found that all parameters constitute that ratio approximately 15%–60% which suggest that the parameters are spatially dependent and are suitable for spatial prediction. Interpolation map of TDS (Fig. 5) highlights the risk zone shown in sky blue, yellow, and red colors. Since tolerable limit of TDS as settled by WHO is 1,000 mg/L; therefore, major part of study area is exceeding the acceptable limit. Concentration of magnesium has also been plotted in Fig. 5. Here, the areas shown in yellow and red colors are at high risk. Calcium concentration is alarming in small area where it is violating the permissible limit (75 mg/L).

Tolerable limit of chloride concentration is 250 mg/L whereas it varies between 11 and 4,255 mg/L in our sample data. As shown in Fig. 5, the major part of observed surface have tolerable chloride concentrations except a small area falling between east latitude 31°–31.2° and north longitude 72.5°–72.8°. Spatial interpolation map of Na^{+} is also shown

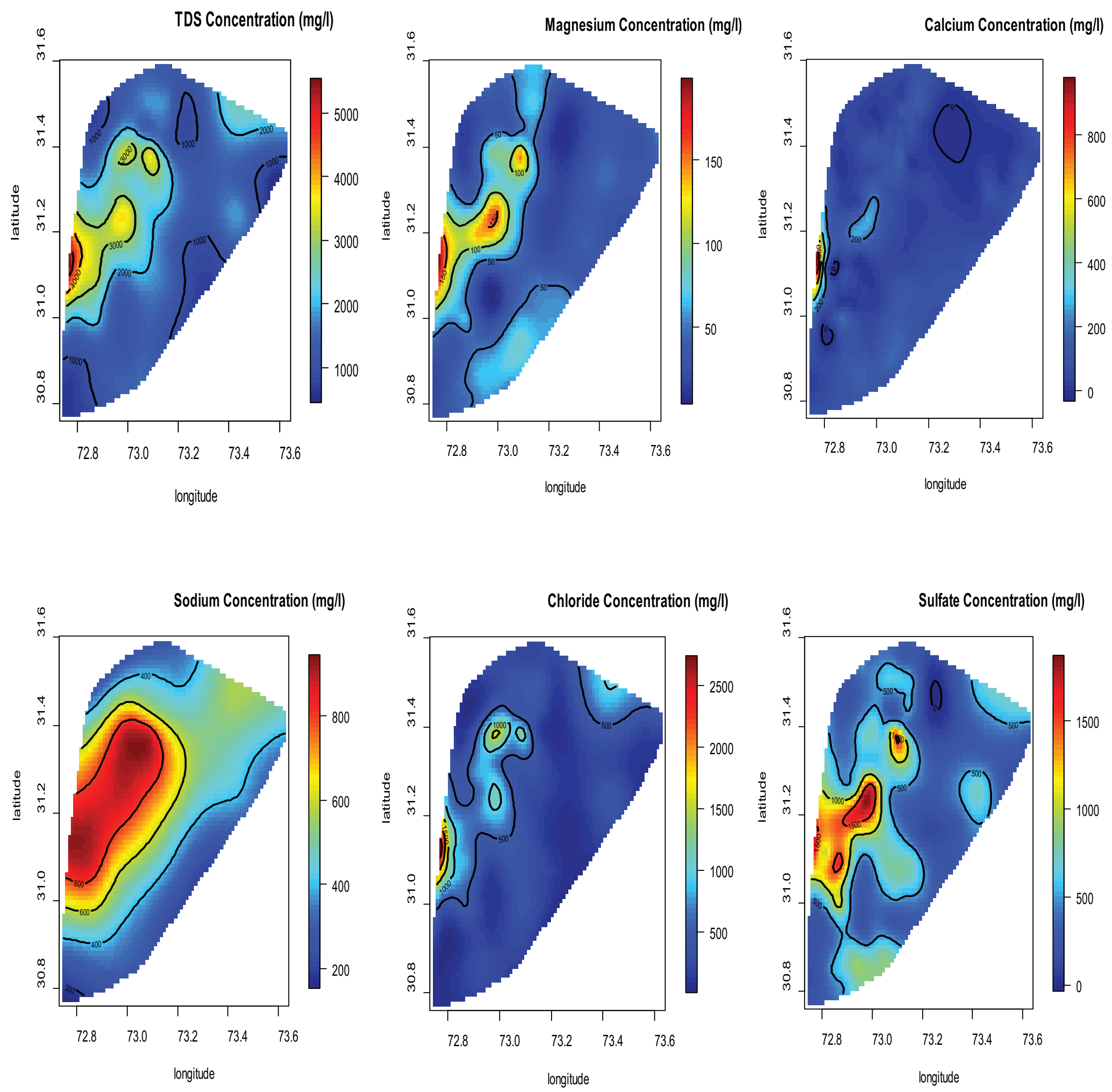

Fig. 5. Spatial prediction maps of water quality parameters: TDS, magnesium, calcium, sodium, chloride, and sulfate, generated using cokriging method.

in Fig. 5. The tolerable limit of Na^+ is 200 mg/L where our data ranges from 3 to 2,000 mg/L. Map shows that only small area represented by blue color has acceptable level of sodium while remaining area have high sodium contamination. Prediction map of sulfate concentration is also represented in Fig. 5. Likewise, sulfate concentration in our data ranges from 12 to 2,416 mg/L while WHO described its acceptable limit as 250 mg/L. Only small part of studied region shows acceptable sulfate concentration level.

3.5. Performance of predicted errors

In this research, we used leave-one-out cross-validation method to assess the performance of cokriging and ordinary kriging. Every given observation was estimated by the neighboring values except itself [1]. The mean squared standard error (MSSE), mean error (ME), and mean squared error (MSE) were used to assess the accuracy of predicted errors during the cross-validation process for both cokriging and ordinary kriging. These measures are defined as follows:

$$\text{MSSE} = \frac{1}{n}\Sigma_{i=1}^{n}\left[\frac{(Y_i^* - Y_i)^2}{\sigma_i^2}\right] \tag{6}$$

$$\text{MSE} = \frac{1}{n}\Sigma_{i=1}^{n}\,(Y_i^* - Y_i)^2 \tag{7}$$

$$\text{ME} = \frac{1}{n}\Sigma_{i=1}^{n}\,(Y_i^* - Y_i) \tag{8}$$

where Y_i and Y_i^* are observed and estimated values, respectively, of given parameters at location *i* and σ_i^2 is the variance of prediction error.

The results of cross-validation indicate better performance of cokriging as compared with ordinary kriging for estimation. Therefore, we recommend estimation of unobserved location using cokriging when multivariate spatial data is under consideration.

Our results showed much consistency with the findings of Rafique et al. [34]. Although their research is based only on descriptive analysis while we illustrated the multivariate spatial aspect of the groundwater parameters as discussed by Triki et al. [1].

4. Conclusion

The results of this study indicate that among 12 groundwater quality parameters, 6 parameters such as TDS, magnesium, calcium, sodium, chloride, and sulfate were found most significant. This significance was initially evaluated by multivariate correlation matrix and later on by PCA and CA. Among significant parameters, TDS was taken as auxiliary variable while remaining five were taken as primary variables. By observing high association among water quality parameters, the use of geostatistical cokriging method is recommended. To declare highly contaminated risk zones in Faisalabad, the resulted predicted values of the cokriging have been illustrated by contour plots. We have adopted the methodology of Triki et al. [1] and found many similarities with their findings. Our results indicate that estimation procedure can be improved to great extent using geostatistical cokriging. Furthermore, it has been observed that the concentrations of TDS, magnesium, sulfate, and sodium are alarming in areas with north longitude 72.8°–73.2° and east latitude 31.0°–31.4°. This work forms a basis for government and other policy makers to improve the quality of groundwater for good health care.

Acknowledgments

Authors are thankful to Mr. Zameer Ahmad Somroo and Mr. Munawar Hussain, Pakistan Council of Research in Water Resources (PCRWR), Lahore, Pakistan for their help in accessing and understanding the data on water quality parameters.

References

[1] I. Triki, N. Trabelsi, M. Zairi, H. Ben Dhia, Multivariate statistical and geostatistical techniques for assessing groundwater salinization in Sfax, a coastal region of eastern Tunisia, Desal. Wat. Treat., 52 (2014) 1980–1989.

[2] J.A. Johnson, C.H. Stoskopf, Comparative Health Systems: Global Perspective, Jones and Bartlett Publishers, LLC, UK, 2009.

[3] UNEP, Groundwater and it as Susceptibility to Degradation: A Global Assessment of the Problem and Options for Management, 2003. Available at: www.unep.org/dewa/Portals/67/pdf/ Groundwater_INC_cover.pdf (Accessed 14 December 2015).

[4] WWF Pakistan, Pakistan's Waters at Risk; Water & Health Related Issues in Pakistan & Key Recommendations, 2007. Available at: www.ircwash.org/sites/default/files/WWF-Pakistan-2007-Pakistans.pdf (Accessed 5 May 2016).

[5] ADB, Asian Water Development Outlook 2013; Measuring Water Security in Asia and Pacific, Asian Development Bank, Philippines, 2013. Available at: https://www.adb.org/sites/default/files/.../asian-water-development-outlook-2013.pdf (Accessed 14 August 2016).

[6] S. Haydar, M. Arshad, J.A. Aziz, Evaluation of drinking water quality in urban areas of Pakistan: a case study of southern Lahore, Pak. J. Eng. Appl., 5 (2009) 16–23.

[7] M.A. Kahlown, M.A. Tahir, H. Rasheed, National Water Quality Monitoring Programme, Fifth Technical Report, Pakistan Council of Research in Water Resources, Islamabad, Vol. 5, 2007. Available at: http://www.pcrwr.gov.pk/Index.html

[8] Z. Ullah, H. Khan, A. Waseem, Q. Mahmood, U. Farooq, Water quality assessment of the River Kabul at Peshawar, Pakistan: industrial and urban wastewater impacts, J. Water Chem. Technol., 35 (2013) 170–176.

[9] A. Waseem, J. Arshad, F. Iqbal, A. Sajjad, Z. Mehmood, G. Murtaza, Pollution status of Pakistan: a retrospective review on heavy metal contamination of water, soil, and vegetables, Biomed. Res. Int., 2014 (2014) 1–29.

[10] M. Iqbal, R.A. Khera, Adsorption of copper and lead in single and binary metal system onto *Fumaria indica* biomass, Chem. Int., 1 (2015) 157–163.

[11] M. Iqbal, M. Muneer, M. Iqbal, M.I. Kanjal, Tannery wastewater treatment using gamma radiation: Kasur tanneries waste management agency, Pakistan, Phys. Chem., 17 (2015) 41–46.

[12] A. Astel, M. Biziuk, A. Przyjazny, J. Namiesnik, Chemometrics in monitoring spatial and temporal variations in drinking water quality, Water Res., 40 (2006) 1706–1716.

[13] F. Sânchez-martos, Mapping groundwater quality variables using PCA and geostatistics: a case study of Bajo Andarax, southeastern Spain, Hydrol. Sci. J., 46 (2001) 227–242.

[14] M.L. Wu, Y.S. Wang, C.C. Sun, H. Wang, J. De Dong, J.P. Yin, S.H. Han, Identification of coastal water quality by statistical analysis methods, Mar. Pollut. Bull., 60 (2010) 852–860.

[15] F. Zhou, Y. Liu, H. Guo, Application of multivariate statistical methods to water quality assessment of the watercourses in northwestern new territories, Hong Kong, Environ. Monit Assess., 132 (2007) 1–13.

[16] M. Yeh, Y. Lin, L.-C. Chang, Designing an optimal multivariate geostatistical groundwater quality monitoring network using factorial kriging and genetic algorithms, Environ. Geol., 50 (2006) 101–121.

[17] B. Nas, Geostatistical approach to assessment of spatial distribution of groundwater quality, Pol. J. Environ. Stud., 18 (2009)1073–1082.

[18] M. Ahmad, S. Chand, Spatial distribution of TDS in drinking water of Tehsil Jampur using Ordinary and Bayesian Kriging, Pak. J. Stat. Oper. Res., 11 (2015) 377–386.

[19] M. Ahmad, S. Chand, H.M. Rafique, Predicting the spatial distribution of sulfate concentration in groundwater of Jampur-Pakistan using geostatistical methods, Desal. Wat. Treat., 57 (2006) 28195–28204.

[20] P.J. Diggle, P.J. Ribeiro, Model-Based Geostatistics, Springer Series in Statistics, Springer, New York, 2007.

[21] L. Pozdnyakova, R. Zhang, Geostatistical Analyses of Soil Salinity in a Large Field, Precis. Agric., 1 (1999) 153–165.

[22] R.T. Mehrjardi, M.Z. Jahromi, S. Mahmodi, A. Heidari, Spatial distribution of groundwater quality with geostatistics (case study: Yazd-Ardakan Plain), World Appl. Sci. J., 4 (2008) 9–17.

[23] Major Cities-Population, Pakistan Demographics Profile 2016, Index Mundi, 2016. Available at: http://www.indexmundi.com/pakistan/demographics_profile.html (Accessed 14 May 2016).

[24] WHO, Guidelines of Drinking Water Quality Recommendation, 3rd ed., World Health Organisation, Geneva, 2006.

[25] A. Rafael Huizar, G. Teodoro Mendez, R. Rafael Madrid, Patterns of groundwater hydrochemistry in Apan-Tochac sub-basin, Mexico, Hydrol. Sci. J., 43 (1998) 669–686.

[26] S. Chandu, N.V. Subbarao, S.N. Chandu, Characterization of groundwater contamination using factor analysis, Environ. Geol., 28 (1996) 175–180.

[27] A.M. Subyani, A.M. Al-Dakheel, Multivariate geostatistical methods of mean annual and seasonal rainfall in southwest Saudi Arabia, Arabian J. Geosci., 2 (2009) 19–27.

[28] R Core Team, A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2014. Available at: http://www.r-project.org/ (Accessed 12 May 2016).

[29] S. Rouhani, Geostatistical Estimation: Kriging, Geostatistics for Environmental and Geotechnical Applications, ASTM International, 1996.

[30] E. Pebesma, B. Graler, Introduction to Spatio-Temporal Variography, *ifigi*, Institute for Geoinformatics, University of Munster. Available at: https://cran.r-project.org/web/packages/gstat/vignettes/st.pdf (Accessed 5 December 2015).

[31] P.J. Ribeiro, P.J. Diggle, geoR: a package for geostatistical analysis, R-NEWS, 1 (2006) 1–28.

[32] G.E. Box, D.R. Cox, An analysis of transformations, J. R. Stat. Soc. Ser. B, Stat. Methods, 48 (1964) 211–252.

[33] J. Fox, S. Weisberg, Package 'Car,' 2016. Available at: https://r-forge.r-project.org/ projects/car/ (Accessed 9 May 2016).

[34] H.M. Rafique, I. Abbas, M.A. Sohl, R. Shehzadi, S.M. Ramay, M. Imran, Y. Al-Zaghayer, A. Mahmood, M.N. Sohl, Appraisal of drinking water quality of tehsil Jampur, Pakistan, Desal. Wat. Treat., 52 (2014) 4641–4648.